

Взаимная энтропия вычисляется последовательным суммированием по строкам (или по столбцам) всех вероятностей матрицы, умноженных на их логарифм:

$$H(AB) = - \sum_i \sum_j p(a_i b_j) \log p(a_i b_j).$$

Путём несложных преобразований также получаем

$$H(AB) = H(A) + H(B | A) = H(B) + H(A | B).$$

СПИСОК ЛИТЕРАТУРЫ

1. Thomsen E. OLAP Solutions: Building Multidimensional Information Systems Second Edition. Wiley Computer Publishing John Wiley & Sons, Inc., 2002.
2. Laks V.S. Lakshmanan, Jian Peiz, Yan Zhao. Qctrees: An efficient summary structure for semantic OLAP. In: SIGMOD, 2003.
3. Мордвинов В.А. Онтология моделирования и проектирования семантических информационных систем и порталов: Справочное пособие. — М.: МИРЭА, 2005. — 237 с.
4. Laks V.S. Lakshmanan, Jian Peiz, Yan Zhao. Socqet: Semantic OLAP with compressed cube and summarization. In: SIGMOD, 2003.
5. Xiaolei Li, Dong Xin Jiawei, Benjamin W. Wah. Star-cubing: Computing iceberg cubes by top-down and bottom-up integration. In: VLDB, 2003.
6. Иванников А.Д., Кулагин В.П., Мордвинов В.А., Найханова Л.В., Овезов Б.Б., Тихонов А.Н., Цветков В.Я. Получение знаний для формирования информационных образовательных ресурсов. — М.: ФГУ ГНИИ ИТТ «Информика», 2008. — 440 с.
7. Yan Zhao. Quotient Cube and QC-Tree: Efficient Summarizations for Semantic OLAP. 2003.

Поступила 25.01.2010 г.

УДК 004.822

ВОПРОСЫ МАТЕМАТИЧЕСКОГО ОБЕСПЕЧЕНИЯ КОМПЬЮТЕРНЫХ СЕТЕЙ НА ОСНОВЕ СЕМАНТИЧЕСКОГО ПОДХОДА

А.Ю. Войтович, В.А. Мордвинов

Московский государственный институт радиотехники, электроники и автоматики (технический университет)

E-mail: voytovich@mirea.ru

Рассматривается обновление подходов к моделированию и формализации описания информационного морфизма на межслойных переходах семантической структуры при сохранении гарантированных показателей пертинентности, релевантности и когнитивности информационных систем, построенных на основе расслоенных архитектур.

Ключевые слова:

Семантическая сеть, информационный морфизм, «сэндвич» Бернерса-Ли, весовой коэффициент, межуровневые переходы, позиции семантико-энтропийное регулирование.

Key words:

Semantic network, information morphism, «sandwich» of Berners-Lee, weight coefficient, interlevel transitions, semantic-entropy control.

Согласно концепции Тима Бернерса-Ли семантическая сеть является интегральным понятием, объединяющим ряд наиболее перспективных направлений развития Интернета и представляет собой многослойную архитектуру, каждый последующий уровень которой отвечает за более тонкие механизмы представления и обработки данных и знаний [1]. Эта структура, рисунок, получила в научных источниках название «сэндвича» Бернерса-Ли:

В основе концепции лежат стандарты, спецификации и рекомендации консорциума W3C (World Wide Web Consortium) для расширяемого

Взаимная энтропия обладает свойством информационной полноты — из неё можно получить все рассматриваемые величины.

Достижимые в результате моделирования полнота вероятностей событий и информационная полнота дают основания применять подходы и результаты показанного здесь моделирования к достаточно широкому спектру разновидностей хО-LAP, в том числе FSeOLAP, TransSeOLAP и другим, представленным в альбоме классификаций OLAP.

языка разметки XML, языка описания структуры XML документа XML Schema, языка запросов XQuery, словаря RDF, унифицированного идентификатора ресурса URI, языка онтологии сетевых сервисов OWL-S [3]. Прогресс концепции заключается в том, чтобы последовательно стандартизировать «семантический сэндвич» — слой за слоем, снизу вверх. Таким образом, выстраиваются механизмы функционирования на каждом уровне. Необходимые средства для формализации информационных процессов на каждом отдельно взятом уровне уже выработаны — их в должном количестве

предоставляют математическая логика и искусственный интеллект, теории формальных и информационных систем [2].



Рисунок. Интерпретация архитектуры семантической сети «сэндвичем» Бернерса-Ли

Вопрос в том, как из имеющегося арсенала выбрать такие методы и средства, которые позволят удовлетворить противоречивые требования к математическому аппарату, возникающие при решении задач моделирования межслоевого взаимодействия. Итак, формализмы должны быть:

- строго определенными на заданном множестве, в том числе на бесконечных множествах;
- достаточно выразительными, чтобы иметь средства для решения нетривиальных задач;
- легкими с алгоритмической точки зрения, чтобы обеспечивать решение задач за допустимое время;
- интуитивно понятными для сохранения главного преимущества семантических сетей в модели «сэндвича» Бернерса-Ли — машинного понимания содержания контента системы.

При таких условиях построение информационной среды, реализующей принципы «семантического сэндвича», является задачей нетривиальной.

В настоящее время в литературе, находящейся в открытом доступе, отсутствуют специализированные методы формализации рассмотренной задачи моделирования семантических информационных сетей и систем, построенных на их основе. В связи с этим, авторам видится целесообразным, учитывая специфику решаемой задачи воспользоваться методами синергетической теории информации для её решения.

С позиции синергетики носитель информации возникает в результате самопроизвольного нарушения существующей симметрии информационного морфизма в точке бифуркации как следствие синергетического развития информационного объекта. Под информационным морфизмом в данной статье понимается взаимодействие, представляющее протяженный во времени процесс взаимозависимого изменения параметров состояния информационного объекта и информационного пространства [4]. В процессе этого взаимодействия в информационной системе объект-источник не всегда испытывает эмиссию, то есть не всегда теря-

ет некоторое количество информации при передаче ее другому объекту, в то время как другой объект-получатель всегда приобретает некое новое добавочное количество информации. Совокупный объем информации, а также их суммарная энтропия при этом обмене неизбежно возрастают, а совокупная система расширяется. Процесс этот ассиметричен.

Композиции информационных морфизмов в системе должны удовлетворять условиям ассоциативности, а для каждого элемента композиции должен быть определен тождественный морфизм. При этом морфизмы двойственной или n -арной категории могут возникать в результате транспортирования матричных форм и перехода к двойственным или множественным отображениям для каждого элемента.

Исходя из этих представлений, авторами гипотезируется существование информационного морфизма, определенного на пространстве моноидов. В литературе информационный морфизм определяется, как гомоморфизм свободного моноида в информационном поле [5].

Вероятностная модель информационного морфизма V_i взаимодействия двух объектов A и B в информационной среде определяется следующим образом:

$$V_i = \frac{C_i}{(E_a + fE_b)}, \quad (*)$$

где C_i — относительное количество информации вида $I_{[k-(k+1)]}$ в дуплексном информационном пространстве (самый общий случай информационного обмена между объектами A и B), где k — максимально возможное количество уровней у одновентной вертикали архитектуры семантической сети; E_a и E_b — относительные (долевые) распределения информации в потоках; f — сложный коэффициент, который определяется как мера энтропийно-семантического соответствия через коэффициент точности информационного обеспечения или коэффициенты полноты, шума, соответствия (в зависимости от специфики поставленных задач).

Коэффициент f в первом приближении равен натуральному числу e в степени произведения:

$$-L(\vec{G}_{ai} + \vec{G}_{bi}),$$

где L — коэффициент Лагранжа, \vec{G}_{ai} и \vec{G}_{bi} — характеристические коэффициенты информационных потоков в направлениях от A к B и от B к A в декартовой системе координат.

Однако, рассмотренного определения недостаточно для представления информационных процессов, происходящих с элементами, одновременно взаимодействующими внутри собственного слоя и в межслойном пространстве, в том числе для открытых информационных систем.

В связи с этим, авторами предлагается использовать модификацию формулы (*), заключающуюся

юся в определении нового характеристического коэффициента информационного потока \vec{G}_{ci} , распространяющегося в произвольном направлении, не совпадающим с плоскостью, на которой описаны внутрислойные взаимодействия.

Тогда

$$f = e^{-L \cdot (\vec{G}_{ai} + \vec{G}_{bi} + \vec{G}_{ci})}.$$

Вероятностная модель информационного морфизма V_i , будет иметь вид:

$$V_i = \frac{C_i}{(E_a + E_b \cdot e^{[-L \cdot (\vec{G}_{ai} + \vec{G}_{bi} + \vec{G}_{ci})]})}.$$

Следовательно, математическое описание информационных морфизмов на переходах из слоя в слой в представлении семантической сети в виде многоуровневого «сэндвича» Бернерса-Ли, имеет следующую форму записи:

$$\begin{aligned} V_i^{[(k-(k+1))]} &= \frac{C_i^{[(k-(k+1))]}{(E_k + f^{[(k-(k+1))]} E_{k+1})} = \\ &= \frac{C_i^{[(k-(k+1))]}{(E_k + (e^{(-L \cdot (G_{ai} + G_{bi} + G_{ci}) \cdot [k-(k+1)])}) E_{k+1})}. \end{aligned}$$

На основе эмпирических результатов формулируется гипотеза, отражающая следующее свойство информационного морфизма расслоенных семантических сетей: «Межуровневые переходы в ком-

плексной модели информационного морфизма в расслоенной архитектуре с прикладными нижними уровнями осуществляют неоднородный вклад в оценку обобщенного морфизма системы (сети)».

Обобщение приведенного выше математического описания выглядит следующим образом:

$$\sum_{k=1}^m V_i^{[(k-(k+1))]} = Z_n \frac{\sum_{i=1}^m C_i^{[(k-(k+1))]}{(E_k + f^{[(k-(k+1))]} E_{k+1})},$$

где m – число уровней; $Z_n = [0:1]$ – весовой коэффициент, отражающий передачу свойств объекта при межуровневом информационном взаимодействии в расслоенной системе; $Z_n = 0$ – незначимый морфизм; $Z_n = 1$ – значимый морфизм.

Выводы

Предложен метод формализации описания межслойных взаимодействий в семантических информационных сетях, основывающийся на модификации обобщенной формулы информационного морфизма информационных объектов. Полученные результаты позволяют применять методы семантико-энтропийного регулирования информационного морфизма к моделированию функционала информационных систем, построенных на основе расслоенных архитектур.

СПИСОК ЛИТЕРАТУРЫ

1. Berners-Lee T., Hendler J., Lassila O. The Semantic Web: Overview / Semantic Web // The Scientific American. – 2008. – № 5. – P. 32–45.
2. Hitzler P., Krotzsch M., Rudolph S. Foundations of Semantic Web Technologies. – U.S.: Chapman & Hall/CRC, 2009. – 455 p.
3. De Bruijn J., Fensel D., Kerrigan M., Keller U., Lausen H., Scicluna J. Modeling Semantic Web Services: The Web Service Modeling Language. – Germany: Springer 2008. – 192 p.
4. Иванников А.Д., Кулагин В.П., Миронов А.А., Мордвинов В.А., Сигов А.С., Тихонов А.Н. Синергетическая теория информационных процессов и систем / под ред. д.пед.н., проф. А.Б. Фоминой. – М.: МГДД(Ю)Т, МИРЭА, ФГУ ГНИИ ИТТ «Информика», 2009. – 550 с.
5. Иванников А.Д., Кулагин В.П., Мордвинов В.А., Найханова Л.В., Овезов Б.Б., Тихонов А.Н., Цветков В.Я. Получение знаний для формирования информационных образовательных ресурсов. – М.: ФГУ ГНИИ ИТТ «Информика», 2008. – 440 с.

Поступила 25.01.2010 г.